Nanopore sequencing accurately identifies the mutagenic DNA lesion O^6-carboxymethyl guanine and reveals its behavior in replication

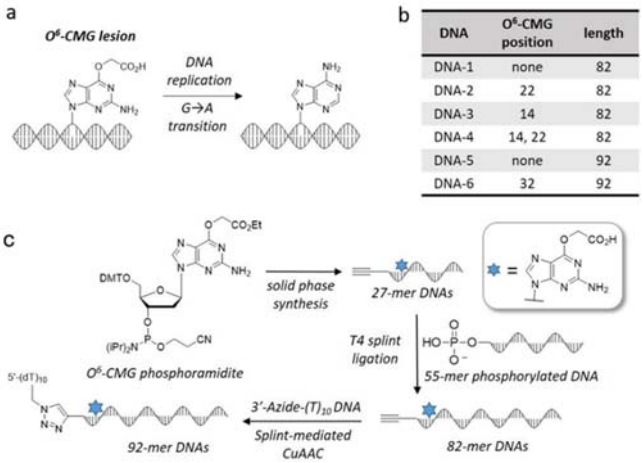
Yu Wang,[#[a]] Kiran M. Patil,[#[b]] Shuanghong Yan,[[a]] Panke Zhang,[[a]] Weiming Guo,[[a]] Yuqin Wang,[[a]] Hong-Yuan Chen,[[a]] Dennis Gillingham,[*[b]] Shuo Huang[*[a]]

[#] These authors contributed equally to this work

Abstract: O^6-carboxymethylguanine (O^6-CMG) is a highly mutagenic alkylation product of DNA, triggering transition mutations relevant to gastrointestinal cancer. However, precise localization of a single O^6-CMG with conventional sequencing platforms is challenging. Here nanopore sequencing, which directly senses single DNA bases according to their physiochemical properties, was employed to detect O^6-CMG. A unique O^6-CMG signal was observed during nanopore sequencing and a single-event call accuracy of >95% was achieved. Moreover, O^6-CMG was found to be a replication obstacle for Phi29 DNA polymerase (Phi29 DNAP), suggesting this lesion could cause DNA sequencing biases in next generation sequencing approaches.

Depurinations,[1] base modifications[2] or DNA strand breaks[3] occur regularly in the genome as a result of exposure to endogenous or exogenous agents.[4] These can interfere with the high-fidelity transmission of genetic information[5] or lead directly to cell death[6] if they are not repaired. Alkylated DNA adducts are often carcinogenic,[7] and are formed by the attack of electrophilic alkylating agents on the nitrogen or oxygen atoms of DNA bases.[8] While attack by these agents is thought to have low sequence specificity, recent efforts have shown that it is the selectivity of DNA repair rather than the selectivity of the initial damage that determines mutational hotspots. For example, with cisplatin[2a, 7, 9] and dihydropyrimidine dimers,[10] densely packed heterochromatic regions and transcription factor binding sites are most prone to mutation.[11] Studies of this sort highlight the need for single nucleotide resolution in sequencing DNA lesions.

O^6-alkyl adducts at guanine are potent mutagens,[12] which occur naturally[13] or during chemotherapy.[14] O^6-Carboxymethylguanine (O^6-CMG, Fig. 1a) is an important O^{6}G lesion because it shows a high tendency to cause miscoding in DNA replication.[15] Indeed its appearance *in vivo* is proportional to the consumption of red meat and may lead to colorectal cancer.[16] Precise detection of O^6-CMG within the genome is important to unveil the mechanisms of its formation, to determine whether certain sites are resistant to repair,[11a] or to explore its potential as a tumor biomarker.[17] O^6-CMG has only been detected by immunoaffinity-HPLC fluorescence assay or LC-MS/MS,[9b] which require a lot of sample,[18] a poorly available antibody,[19] or lose sequence information because of strand digestion.[20] Hence, no current method can directly localize individual O^6-CMG bases within a stretch of DNA. Nanopore sequencing (NPS)[21] would seem to be the ideal platform for characterizing DNA lesions because it achieves base-calling by directly measuring a property of the molecule; additionally NPS is amplification-free, has single-nucleotide resolution, and can read non-canonical bases.[22]



DNA	O^6-CMG position	length
DNA-1	none	82
DNA-2	22	82
DNA-3	14	82
DNA-4	14, 22	82
DNA-5	none	92
DNA-6	32	92

Figure 1. **Structure and synthesis of DNA with O^6-CMG lesions. a)** The O^6-CMG lesion occurs in DNA and is a potent mutagen, leading to G→A transition mutations. **b)** Six DNA strands bearing zero to two O^6-CMG lesions at different positions were synthesized. **c)** Solid-phase oligonucleotide synthesis with a modified phosphoramidite and one (82-mer DNAs) or two (92-mer DNAs) splint mediated ligation steps deliver all of the oligonucleotides for NP testing.

In the present study, NPS was used to identify O^6-CMG lesions in various locations within synthetic DNA strands (Fig. 1c). The lesions were unambiguously spotted according to their characteristic signals in the *Mycobacterium smegmatis* porin A (MspA) nanopore (NP) assisted by Phi29 DNAP (see the electronic supporting information (ESI) for details)[23] with a single event call accuracy of >95.2%. The data also reveals that O^6-CMG is a potent replication obstacle for Phi29 DNAP both at the single molecule level and in ensemble. Our results demonstrate

[a] Ms. Y. Wang, Dr. S. H. Yan, Bc. W. M. Guo, Dr. Y. Q. Wang, Dr. P. K. Zhang, Prof. H. Y. Chen, Prof. S. Huang
School of Chemistry and Chemical Engineering
State Key Laboratory of Analytical Chemistry for Life Science and Collaborative Innovation Center of Chemistry for Life Sciences
Nanjing University, 210023, China
E-mail: shuo.huang@nju.edu.cn

[b] Dr. K. M. Patil, Prof. D. Gillingham
Department of Chemistry
University of Basel, CH-4056, Basel, Switzerland
E-mail: dennis.gillingham@unibas.ch

Supporting information for this article is given via a link at the end of the document.

the first sequencing strategy for O^6-CMG while also highlighting the potential of single molecule studies for understanding enzymatic activities.

Sequence information from MspA NPS is inferred by reading different combinations of sequence quadromers simultaneously occupying the pore restriction.[23] According to the published quadromer map,[23] AGAA and CTTT report the highest and the lowest current level respectively as they pass through the MspA NP. According to the enzymatic ratcheting mechanism of NPS,[22a] sequential NP reading of a CTTTAGAACTTT sequence segment yields an asymmetric triangle-shaped current signature with nine successive current steps cycling between the lowest (CTTT) and highest (AGAA) signals. In this work, DNA templates to be sequenced were designed to contain four of these highly distinguishable AGAACTTT cycles (Fig. 1b, Table S1).

To evaluate the NP sensing ability of O^6-CMG, a set of oligonucleotides bearing the lesion at different sites were prepared. The syntheses began with a phosphoramidite prepared by a copper-catalyzed O^6G alkylation with commercially available ethyldiazoacetate, as described previously (Fig. S1, S2).[24] With this phosphoramidite, O^6-CMG bearing 27-mer DNAs with a 5'-alkyne tag were prepared by solid-phase synthesis (Fig. S3) and then coupled to a 5'-phosphorylated canonical 55-mer DNA using a T4 splint ligation (Fig. 1c). The resulting 82-mer DNAs could be further extended to the longer 92-mers containing a deca(deoxythymidine) tag, by performing a splint-mediated Cu-catalyzed azide alkyne cycloaddition. Using this approach, three 82-mers (DNA-2, DNA-3, DNA-4) and one 92-mer DNA template (DNA-6), which contain O^6-CMG modifications, were produced, in addition to the necessary unmodified reference strands (DNA-1 & DNA-5). The 82-mers were used to study the performance of O^6-CMG spotting by NPS while the 92-mers were used for the single molecule studies of Phi29 DNAP.

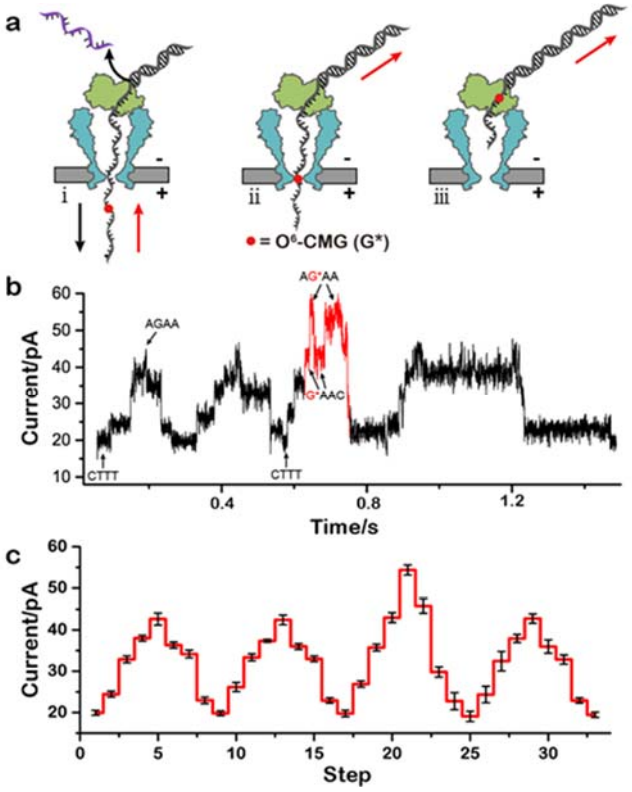


Figure 2. **Direct spotting of O^6-CMG by NPS. a)** Schematic of O^6-CMG reading during NPS. The DNA library is composed of the template strand (which contains the O^6-CMG, red dot), the primer strand and the blocker (purple). In an applied electric potential, the DNA library was electrophoretically driven into the MspA NP (blue). (i) The scenario of voltage-driven unzipping (black arrows) and replication-driven (Phi29 DNAP, green) ratcheting (red arrow) represents the initiation of NPS. (ii) O^6-CMG on the template strand is spotted when it enters the pore constriction. (iii) The analyte escapes from the NP as the end of the template strand is reached. **b)** A representative electrophysiology trace from NPS of DNA-2 (Table S1). The red segment of the trace (flickering) is diagnostic for O^6-CMG in the restriction (Fig. S9). **c)** Statistical signal pattern of DNA-2. The means and standard deviations are from independent sequencing trials (N=20). The higher signal around step 21 is the O^6-CMG.

NPS was carried out using the enzymatic ratcheting strategy[22a] with a mutant MspA NP[22c, 25] (see ESI) inserted in a self-assembled 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) lipid bilayer, which separates the *cis* and the *trans* side of a custom-made electrophysiology chamber.[21a] Wildtype Phi29 DNAP was used in the *cis* side of the chamber to promote the DNA copying for NPS (fully described in the ESI, Fig. S4). The sequencing library is composed of three parts: the DNA template to be sequenced, the blocker for preventing the DNA template from being extended by Phi29 DNAP in solution and the primer for initiating DNA replication (Table S1).[22b] Synthetic ssDNAs DNA-n (n=1-6, Fig. 1b, Table S1) were thermally annealed with the primer and the blocker strands by published protocols (see ESI).[22b] The 5' end of the DNA template was electrophoretically driven into the pore first, thereby releasing the blocker strand (Fig. 2ai).[26] This was followed by the Phi29 DNAP-driven primer extension (Fig. 2aii-iii). NPS signals appear as step transitions during stages ii and iii following a 3'-5' order of signal appearance

until the template-primer complex escapes the pore (Fig. 2aiii). The template strand DNA-1 (Table S1), which is composed only of canonical DNA bases, was taken as a reference strand for NPS. Four consecutive triangle-shaped current segments (derived from the four sequence repeats of AGAACTTT) were consistently monitored from the raw electrophysiology traces (Fig. S5). A zoomed-in view of a single period of the signal shows detailed steps within the sequencing trace, in which AGAA and CTTT report the highest and the lowest level of NP readout, respectively. The asymmetric pattern of the NPS signal inherent to this particular sequence design is helpful to identify abnormal backward motion of the enzyme, which occasionally occurs during NPS.[22a]

NPS of DNA-1 generates signals with high consistency across many independent sequencing events (see Fig. S6 for statistics and normalization procedure). Other DNA-n strands (n=2-6, Table S1), which contain identical sequences but with specific alkylated guanines, are expected to give similar current signals except at locations with O^6-CMG. For example, DNA-2 contains a single O^6-CMG replacement located in the 3rd cycle of the sequence repeats. As expected, four cycles of triangular shaped signals were consistently recorded when DNA-2 was NP sequenced. In contrast to DNA-1, however, an abnormally high and flickering signal was observed around the peak of the 3rd cycle, consistent with the location of the O^6-CMG nucleotide (Table S1, Fig. 2b). This distinctive signal level, which corresponds to NP reading of AG*AA (G*=O^6-CMG) instead of AGAA in the reference (DNA-1, Fig. S5-S6) was observed consistently as demonstrated in statistics from 20 independent NPS reads of DNA-2 (Fig. 2c). After normalization of the traces, the mean amplitude difference between the statistical values acquired from DNA-2 and DNA-1 demonstrates a ~10 pA difference (Fig. S8) around the location of an O^6-CMG in DNA-2. Among all the quadromer combinations of canonical bases, the NP reading of AGAA reports the largest NPS current level.[23] NP reading of AG*AA (G*=O^6-CMG), however, reports a further elevation of 10 pA above the AGAA readout, which clearly differentiates the AG*AA readout from that of all other base combinations. In addition, characteristic trace flickering was recorded whenever AG*AA reached the NP restriction (red segment, Fig. 2b). The trace flickering, which appears as forward and backward step transitions between AG*AA and G*AAC, may result from either dangling motions of the carboxylmethyl group of O^6-CMG or the protonation equilibrium of the acid (Fig. S9). Both the high current step and the flickering signal are unique criteria that can be used to identify O^6-CMG.

To estimate the call accuracy of O^6-CMG in DNA-2, the highest current level, which corresponds to NP reading of AG*AA, was taken as the judgement criterion. There were 42 raw sequencing events in the statistics, and in only two cases did the AG*AA signal fail to reach the expected level, giving a single-event spotting accuracy of 95.2% (Fig. S10). The 4.8% of failed reads may result from either too short an acquisition time for the O^6-CMG occupying the pore restriction or hydrolysis of the carboxymethyl group from the base. Nevertheless, we conclude conservatively that the single-event spotting accuracy is more than 95.2% under the conditions tested.

DNA-3 and DNA-4 (Fig. 1b, Table S1) were sequenced next; these are identical to DNA-1 except for one (DNA-3) or two (DNA-4) O^6-CMG base replacements at different positions. To highlight the robustness of O^6-CMG reading during NPS, the statistical mean differences between DNA-2, 3 and 4 compared to DNA-1 are summarized in Fig. 3. To avoid pore-to-pore variations, current traces were normalized before the derivation of the statistical results. The results (Fig. 3, Fig. S11-S12) show that different quantities or locations of O^6-CMG can be unambiguously spotted by a consistent 10 pA elevated current relative to the reference strand.

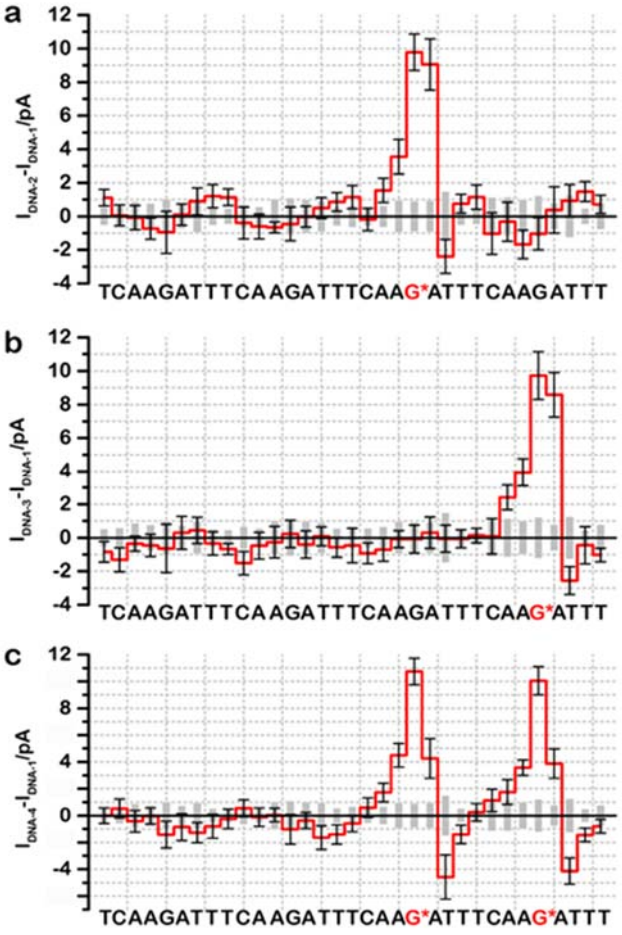


Figure 3. **Mean current differences relative to DNA-1 between DNAs possessing different O^6-CMG sites.** DNA-2, 3, 4 (Table S1), contain different O^6-CMG sites within the sequence. O^6-CMGs in each template are marked as G* in red within the sequence below each chart **(a-c)**. Red lines in each charts show the mean current differences between alkylated DNAs and the reference (DNA-1, Table S1). 20 readouts from each template were used to form the statistics. Black bars on red lines show the standard deviations of alkylated DNAs while the gray columns along the x-axis show the standard deviations from DNA-1. **a)** DNA-2 **b)** DNA-3 **c)** DNA-4. From a-c, a mean current difference of ~10 pA was observed between O^6-CMG and guanine.

Next the interaction of the O^6-CMG lesion with Phi29 DNAP was analyzed using the NP as a molecular tweezer. A NP tweezer is a device where a known DNA sequence serves as a molecular ruler as it is ratcheted out of the pore during sequencing.[27] In the present case, the NPS setup could be used to probe Phi29 DNAP activity as it meets O^6-CMG on the single molecule level with a spatial resolution as low as ~40 pm. The height of the MspA NP requires ~15 nucleotides to span from the NP restriction to bulk solvent, where Phi29 DNAP is located.[28] Although the 82-mer DNAs (DNA2-4 in Fig. 1b and Table S1) would be just long enough to reach out of the pore, the reduced electrophoretic force (fewer charges remaining in the pore) would allow the DNA to escape the pore too quickly by Brownian motion. Hence 92-mer DNAs (DNA-5 and DNA-6, see Fig. 1c for synthesis and Table S1 for precise structures) bearing a 10-thymidine tail installed through a triazole linkage were used in the molecular tweezers experiments. As expected, DNA-5, which contains no O^6-CMG, delivers a nearly identical NPS signal as DNA-1 (Fig. S13). However, due to the added length of DNA at the 5' end of DNA-5, more sequencing events could be recorded (Fig. S13a). All NPS signals acquired from DNA-5 end with spontaneous pore restoration, which means that the added triazole linker and the extra 10-thymidine tag itself cannot hold DNA-5 efficiently in the pore restriction. DNA-6 (Table S1) is identical to DNA-5 except that it bears an O^6-CMG nucleotide replacement at position thirty-two. Hence comparing DNA-5 and DNA-6 should provide information on the single molecule interaction of the O^6-CMG with the Phi29 DNAP (Fig. 4a). During NPS of DNA-6 (Fig. 4b) four consecutive peaks appeared with an elevated signal in the 3rd peak of the trace (first 0.6 seconds), fully consistent with the data of DNA-2. However, this was followed by a halt of the enzymatic activity and a flickering signal a further ~16 steps (*i.e.* nucleotides) beyond the appearance of the O^6-CMG signal (Fig. 4bii' and Fig. 4c for zoom-in). This single molecule enzymatic halt phenomenon, which is observed in 98.3% of the events acquired from DNA-6, is never observed with the control sample DNA-5 (Fig. S14). Hence the single molecule NP tweezers assay implies that the O^6-CMG acts as a replication obstacle for Phi29 DNAP in at least 98% of their encounters. To compare these results with ensemble activity of Phi29 DNAP we performed a polymerase extension assay. To avoid interference from the triazole linker, DNA-1 and DNA-2 were used in the polymerase extension assay. DNA-1, which contains no alkylated DNA bases, is extended to the full-length product, as marked by position **1** (Fig. 4d). DNA-2, which contains a single O^6-CMG nucleotide within its sequence, shows a strong stopping band (marked as position **2** on the gel, Fig. 4d). Combined with the results from the single molecule assay, it is clear that O^6-CMG acts as a replication obstacle for Phi29 DNAP. The fact that most of DNA-2 is fully copied (despite the 98% blocking frequency suggested in the single molecule assay) is likely an indication that Phi29 DNAP repeatedly attempts to copy beyond O^6-CMG over the 20-minute timeline of the experiment.

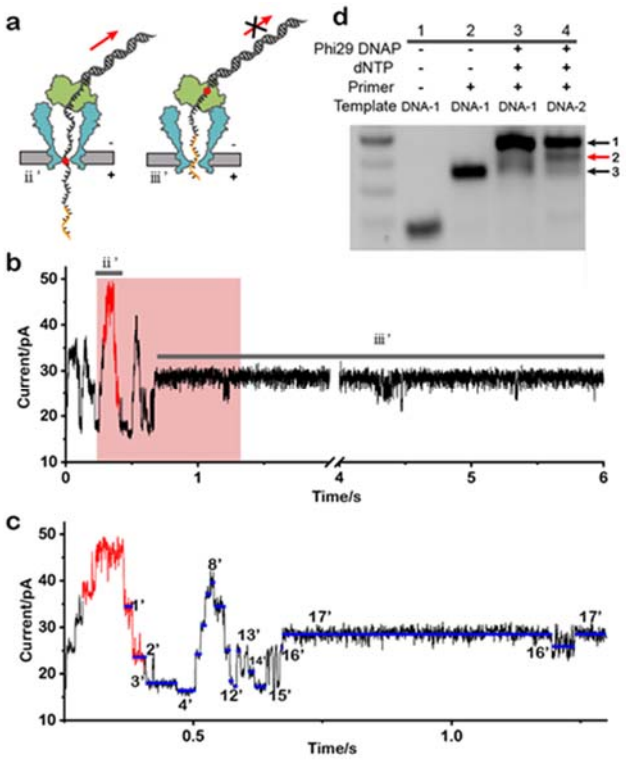


Figure 4. **Single-molecule and ensemble experiments establishing O^6-CMG as a replication obstacle of Phi29 DNAP. a)** Schematic of the molecular tweezer assay. The 92-mer DNA-6 (Table S1) provides enough length for the CMG lesion to encounter Phi29 DNAP. (ii') O^6-CMG (red dot) can be spotted as it translocates through the pore constriction. The red arrow indicates the direction of motion of the DNA template when ratcheted by the Phi29 DNAP during DNA replication. (iii') When O^6-CMG reaches the binding pocket of the Phi29 DNAP, further DNA replication stops. **b)** A representative current trace acquired from DNA-6 during the single molecule assay. Characteristic signals of O^6-CMG (ii', the red segment) were first observed. After subsequent chain elongation of ~14 bases, repetitive fluctuating (iii') without further elongation is observed, indicating that O^6-CMG is an obstacle for Phi29 DNAP. **c)** A zoom in of the first 1.5 seconds of the tweezer assay shows a 16'-17' nucleotide phase-shift between AG*AA reading and the subsequent enzymatic halt (Fig. S14), after which the flickering signal is observed. **d)** A gel assay establishes O^6-CMG as a replication obstacle for Phi29 DNAP in ensemble (conditions in the ESI).

In summary, we have demonstrated that the mutagenic DNA lesion O^6-CMG can be read by direct NPS. Abnormally high current readouts with flickering characteristics were observed when O^6-CMG occupied the pore restriction during NPS. Assisted by these unique signal features, a single-event call accuracy of 95.2% was achieved for O^6-CMG reading. As the first demonstration of direct O^6-CMG detection in a single molecule, this work opens the door for more precise characterization of O^6-CMG in cellular DNA, which at present would be challenging to sequence with single-nucleotide resolution with any NGS platform. The conclusions in this paper suggest that the commercially available MinIon™ sequencer could directly identify O^6-CMG in DNA samples.

Acknowledgements

This work is supported by National Natural Science Foundation of China (Grant No. 21327902, Grant No. 21675083, Grant No. 91753108), Fundamental Research Funds for the Central Universities (Grant No. 020514380142, No. 020514380174), State Key Laboratory of Analytical Chemistry for Life Science (Grant No.5431ZZXM1804, No. 5431ZZXM1902), 1000 Plan Youth Talent Program of China, Program for high-step entrepreneurial and innovative talents introduction of Jiangsu Province,Technology innovation fund program of Nanjing University.

D. G. acknowledges support from the Swiss National Science Foundation (Grant No. 200021_172521).

Keywords: nanopore sequencing • DNA lesion • carboxymethyl guanine • gastrointestinal cancer •

References

[1] T. A. Kunkel, *Proc. Natl. Acad. Sci. U.S.A* **1984**, *81*, 1494-1498.
[2] (a) F. Drablos, E. Feyzi, P. A. Aas, C. B. Vaagbo, B. Kavli, M. S. Bratlie, J. Pena-Diaz, M. Otterlei, G. Slupphaug, H. E. Krokan, *DNA Repair* **2004**, *3*, 1389-1407; (b) P. Dandona, K. Thusu, S. Cook, B. Snyder, J. Makowski, D. Armstrong, T. Nicotera, *Lancet* **1996**, *347*, 444-445; (c) K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298.
[3] K. K. Khanna, S. P. Jackson, *Nature Genet.* **2001**, *27*, 247-254.
[4] S. P. Jackson, J. Bartek, *Nature* **2009**, *461*, 1071-1078.
[5] A. Ciccia, S. J. Elledge, *Mol. Cell* **2010**, *40*, 179-204.
[6] A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz, S. Linn, *Annu. Rev. Biochem.* **2004**, *73*, 39-85.
[7] M. H. Raz, H. R. Dexter, C. L. Millington, B. van Loon, D. M. Williams, S. J. Sturla, *Chem. Res. Tox.* **2016**, *29*, 1493-1503.
[8] (a) N. Shrivastav, D. Y. Li, J. M. Essigmann, *Carcinogenesis* **2010**, *31*, 59-70; (b) D. Gillingham, S. Geigle, O. A. von Lilienfeld, *Chem. Soc. Rev.* **2016**, *45*, 2637-2655.
[9] (a) X. Shu, X. Xiong, J. Song, C. He, C. Yi, *Angew. Chem. Int. Ed.* **2016**, *55*, 14246-14249; (b) Y. Yu, J. S. Wang, P. C. Wang, Y. S. Wang, *Anal. Chem.* **2016**, *88*, 8036-8042; (c) J. Hu, J. D. Lieb, A. Sancar, S. Adar, *Proc. Natl. Acad. Sci. U.S.A* **2016**, *113*, 11507-11512.
[10] J. Hu, O. Adebali, S. Adar, A. Sancar, *Proc. Natl. Acad. Sci. U.S.A* **2017**, *114*, 6758-6763.
[11] (a) D. Gillingham, B. Sauter, *ChemBioChem* **2017**, *18*, 2368-2375; (b) R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, *Nature* **2016**, *532*, 264; (c) D. Perera, R. C. Poulos, A. Shah, D. Beck, J. E. Pimanda, J. W. H. Wong, *Nature* **2016**, *532*, 259.
[12] B. E. Johnson, T. Mazor, C. Hong, M. Barnes, K. Aihara, C. Y. McLean, S. D. Fouse, S. Yamamoto, H. Ueda, K. Tatsuno, *Science* **2014**, *343*, 189-193.
[13] B. C. Cupid, Z. Zeng, R. Singh, D. E. G. Shuker, *Chem. Res. Tox.* **2004**, *17*, 294-300.
[14] J. Zhang, M. F.G. Stevens, T. D. Bradshaw, *Curr. Mol. Pharmacol.* **2012**, *5*, 102-114.
[15] (a) L. A. Wyss, A. Nilforoushan, D. M. Williams, A. Marx, S. J. Sturla, *Nuc. Acids. Res.* **2016**, *44*, 6564-6573; (b) F. Zhang, M. Tsunoda, K. Suzuki, Y. Kikuchi, O. Wilkinson, C. L. Millington, G. P. Margison, D. M. Williams, E. C. Morishita, A. Takenaka, *Nuc. Acids. Res.* **2013**, *41*, 5524-5532.
[16] (a) M. H. Lewin, N. Bailey, T. Bandaletova, R. Bowman, A. J. Cross, J. Pollock, D. E. G. Shuker, S. A. Bingham, *Cancer Res.* **2006**, *66*, 1859-1865; (b) P. Senthong, C. L. Millington, O. J. Wilkinson, A. S. Marriott, A. J. Watson, O. Reamtong, C. E. Eyers, D. M. Williams, G. P. Margison, A. C. Povey, *Nuc. Acids. Res.* **2013**, *41*, 3047-3055.
[17] (a) Y. Mishina, E. M. Duguid, C. He, *Chem. Rev.* **2006**, *106*, 215-232; (b) J. Riedl, Y. Ding, A. M. Fleming, C. J. Burrows, *Nat. Commun.* **2015**, *6*, 11.
[18] B. C. Cupid, Z. T. Zeng, R. Singh, D. E. G. Shuker, *Chem. Res. Tox.* **2004**, *17*, 294-300.
[19] K. L. Harrison, N. Fairhurst, B. C. Challis, D. E. Shuker, *Chem. Res. Tox.* **1997**, *10*, 652-659.
[20] J. Vanden Bussche, S. A. Moore, F. Pasmans, G. G. C. Kuhnle, L. Vanhaecke, *J. Chromatogr. A* **2012**, *1257*, 25-33.
[21] (a) S. Huang, *Chin. Sci. Bull.* **2014**, *59*, 4918-4928; (b) D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. H. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotech.* **2008**, *26*, 1146-1153; (c) D. Deamer, M. Akeson, D. Branton, *Nat. Biotech.* **2016**, *34*, 518-524.
[22] (a) G. M. Cherf, K. R. Lieberman, H. Rashid, C. E. Lam, K. Karplus, M. Akeson, *Nat. Biotech.* **2012**, *30*, 344-348; (b) E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Nat. Biotech.* **2012**, *30*, 349-U174; (c) T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, J. H. Gundlach, *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 20647-20652; (d) W.-W. Li, L. Gong, H. Bayley, *Angew. Chem. Int. Ed.* **2013**, *52*, 4350-4355; (e) A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, J. H. Gundlach, *Proc. Natl. Acad. Sci. U.S.A* **2013**, *110*, 18904-18909; (f) J. Schreiber, Z. L. Wescoe, R. Abu-Shumays, J. T. Vivian, B. Baatar, K. Karplus, M. Akeson, *Proc. Natl. Acad. Sci. U.S.A* **2013**, *110*, 18910-18915; (g) A. E. P. Schibel, N. An, Q. A. Jin, A. M. Fleming, C. J. Burrows, H. S. White, *J. Am. Chem. Soc.* **2010**, *132*, 17992-17995. (h) N. An, A. M. Fleming, H. S. White, C. J. Burrows, *ACS Nano* **2015**, *9*, 4296-4307.
[23] A. H. Laszlo, I. M. Derrington, B. C. Ross, H. Brinkerhoff, A. Adey, I. C. Nova, J. M. Craig, K. W. Langford, J. M. Samson, R. Daza, K. Doering, J. Shendure, J. H. Gundlach, *Nat. Biotech.* **2014**, *32*, 829-833.
[24] (a) M. H. Räz, E. Sandell, K. Patil, D. G. Gillingham, S. J. Sturla, *ACS Chem. Biol.* **2019**, *Just Accepted, DOI: 10.1021/acschembio.8b00802*; (b) S. N. Geigle, L. A. Wyss, S. J. Sturla, D. G. Gillingham, *Chem. Sci.* **2017**, *8*, 499-506.
[25] Y. Q. Wang, S. H. Yan, P. K. Zhang, Z. D. Zeng, D. Zhao, J. X. Wang, H. Y. Chen, S. Huang, *ACS Appl. Mater. Interfaces* **2018**, *10*, 7788-7797.
[26] Z. L. Wescoe, J. Schreiber, M. Akeson, *J. Am. Chem. Soc.* **2014**, *136*, 16582-16587.
[27] I. M. Derrington, J. M. Craig, E. Stava, A. H. Laszlo, B. C. Ross, H. Brinkerhoff, I. C. Nova, K. Doering, B. I. Tickman, M. Ronaghi, J. G. Mandell, K. L. Gunderson, J. H. Gundlach, *Nat. Biotech.* **2015**, *33*, 1073.
[28] S. H. Yan, X. T. Li, P. K. Zhang, Y. Q. Wang, H. Y. Chen, S. Huang, H. Y. Yu, *Chem. Sci.* **2019**, *Just Accepted, DOI: 10.1039/C8SC05228J.*